

Supporting Information

Machine Learning Driven Prediction of Band Alignment Types in 2D Hybrid Perovskites

Eti Mahal,[†] Diptendu Roy,[†] Surya Sekhar Manna,[†] Biswarup Pathak^{†*}

[†]Department of Chemistry, Indian Institute of Technology Indore, Indore 453552, India

*Email: biswarup@iiti.ac.in

Contents

Figure S1: List of ammonium Cations contained in the 2D hybrid perovskite materials selected in this study

Figure S2: DFT calculated density of states plots

Figure S3: Single MX6 unit with four methyl ammonium cations.

Figure S4: Accuracy score vs. number of features plot

Figure S5: Confusion matrix and classification parameters

Figure S6: Confusion matrix for type I_a and type I_b classification

Figure S7: Confusion matrix for type II_a and type II_b classification

Figure S8: Confusion matrix for multi class classification

Text S1: Computational details

Text S2: Feature extraction details

Text S3: Machine learning algorithms

Text S4: Repeated Stratified k-fold cross-validation

Text S5: Classification metrices

Text S6: Model details

Table S1: Selected features

Table S2: Predicted probability of test dataset

Table S3: Optimized hyperparameters and accuracy scores of binary classification between type I_a-I_b and type II_a-II_b and multi class classification between four types (I_a, I_b, II_a, II_b) using logistic regression.

Table S4: Classification metrices for classification between I_a and I_b

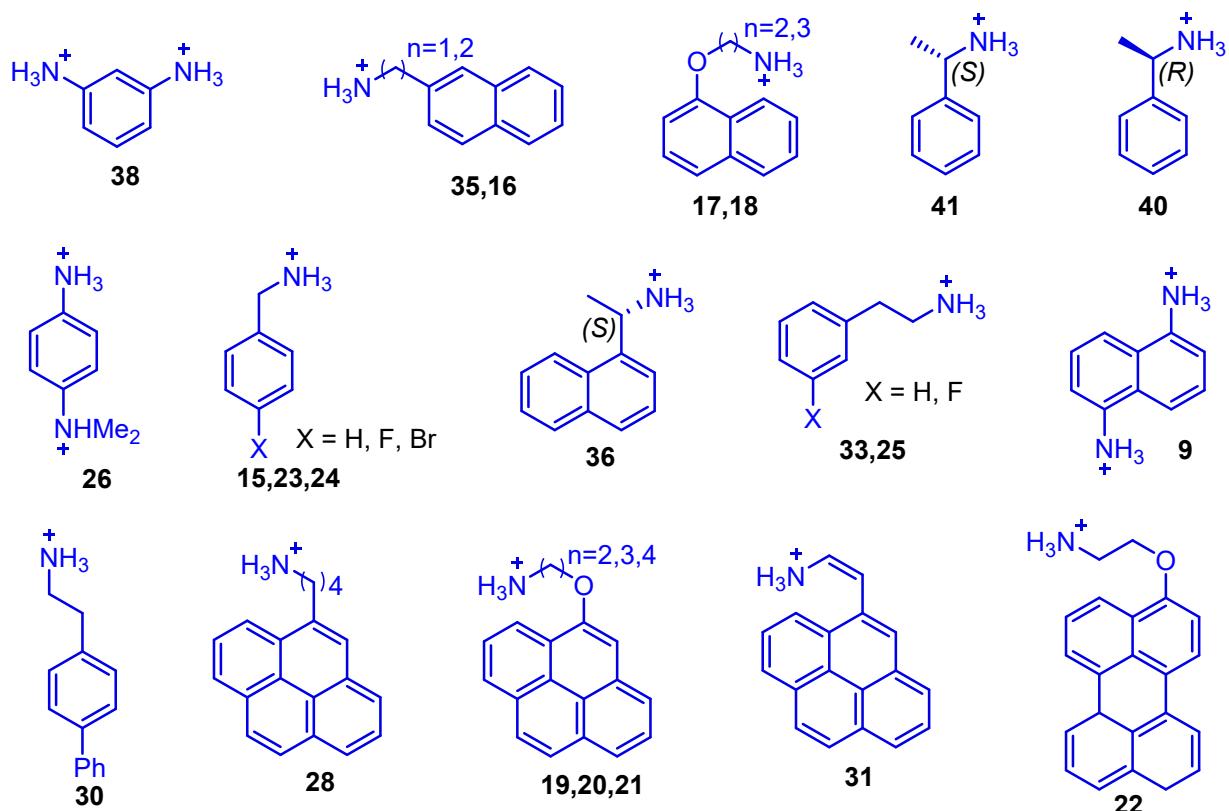
Table S5: Classification metrices for classification between II_a and II_b

Table S6: Classification metrices for multi class classification

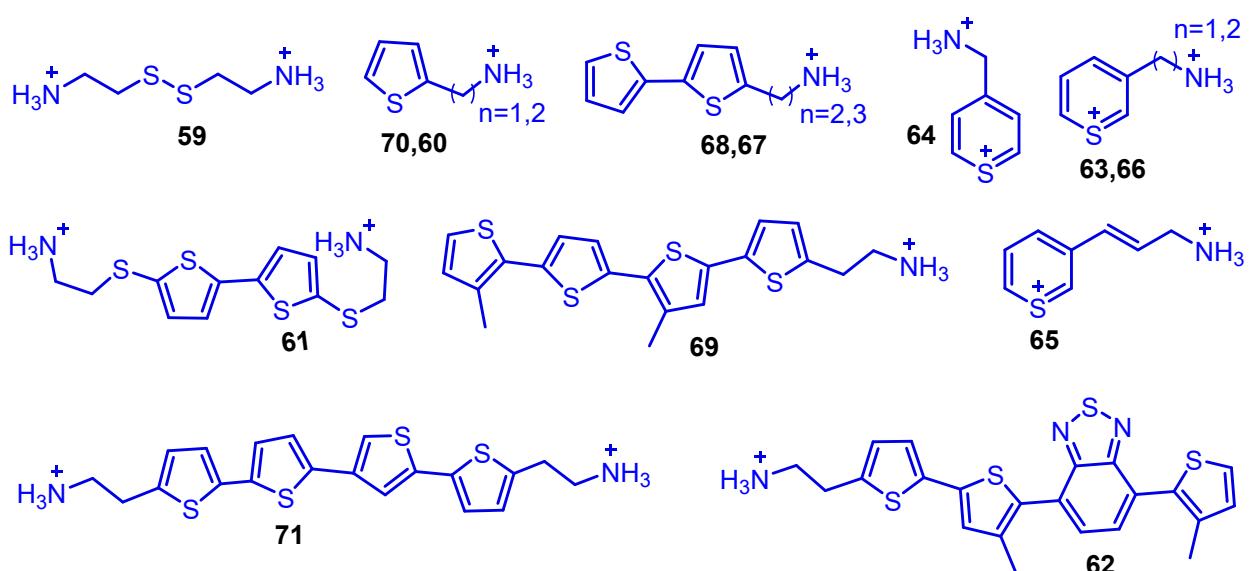
Table S7: DFT validation of predicted unknown dataset

Table S8: Whole dataset of 103 2D perovskite materials considered in this work

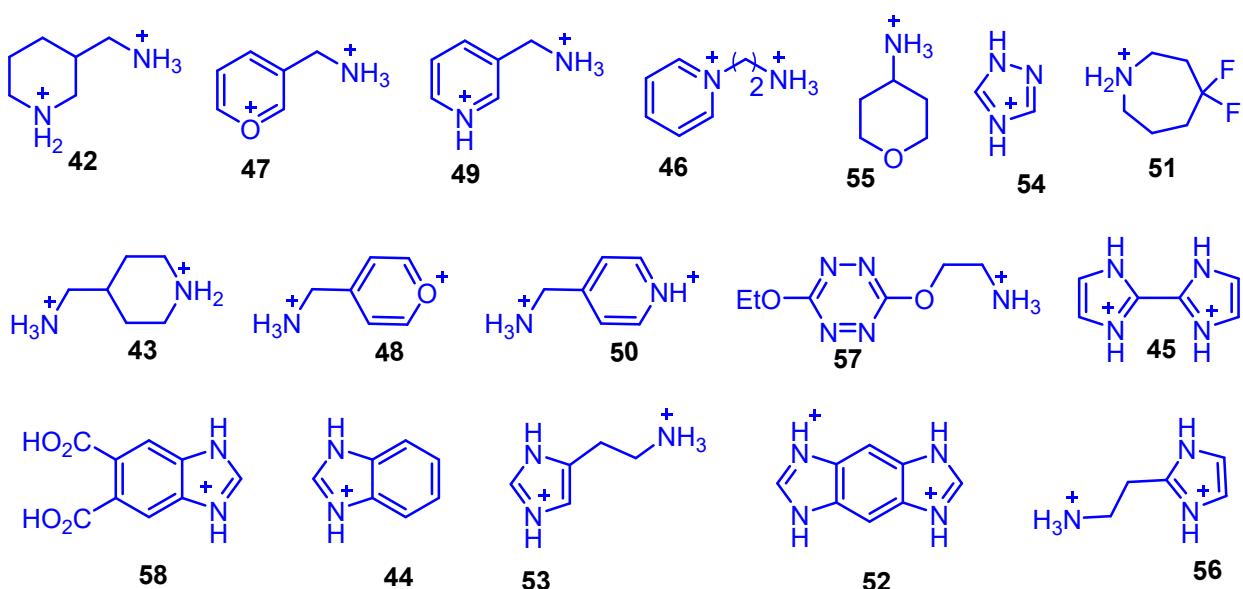
A) Aryl Ring Containing Cations



B) Thiophene and thiopyrylium Ring Containing Cations



C) Heterocyclic Ring Containing Cations



D) Aliphatic Cations

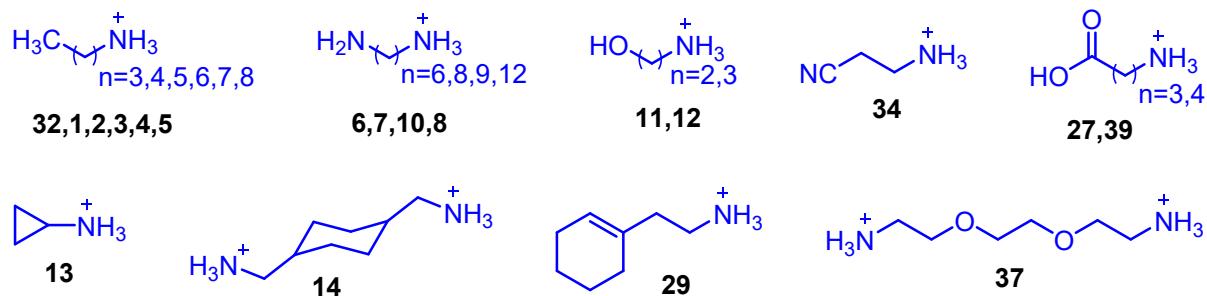
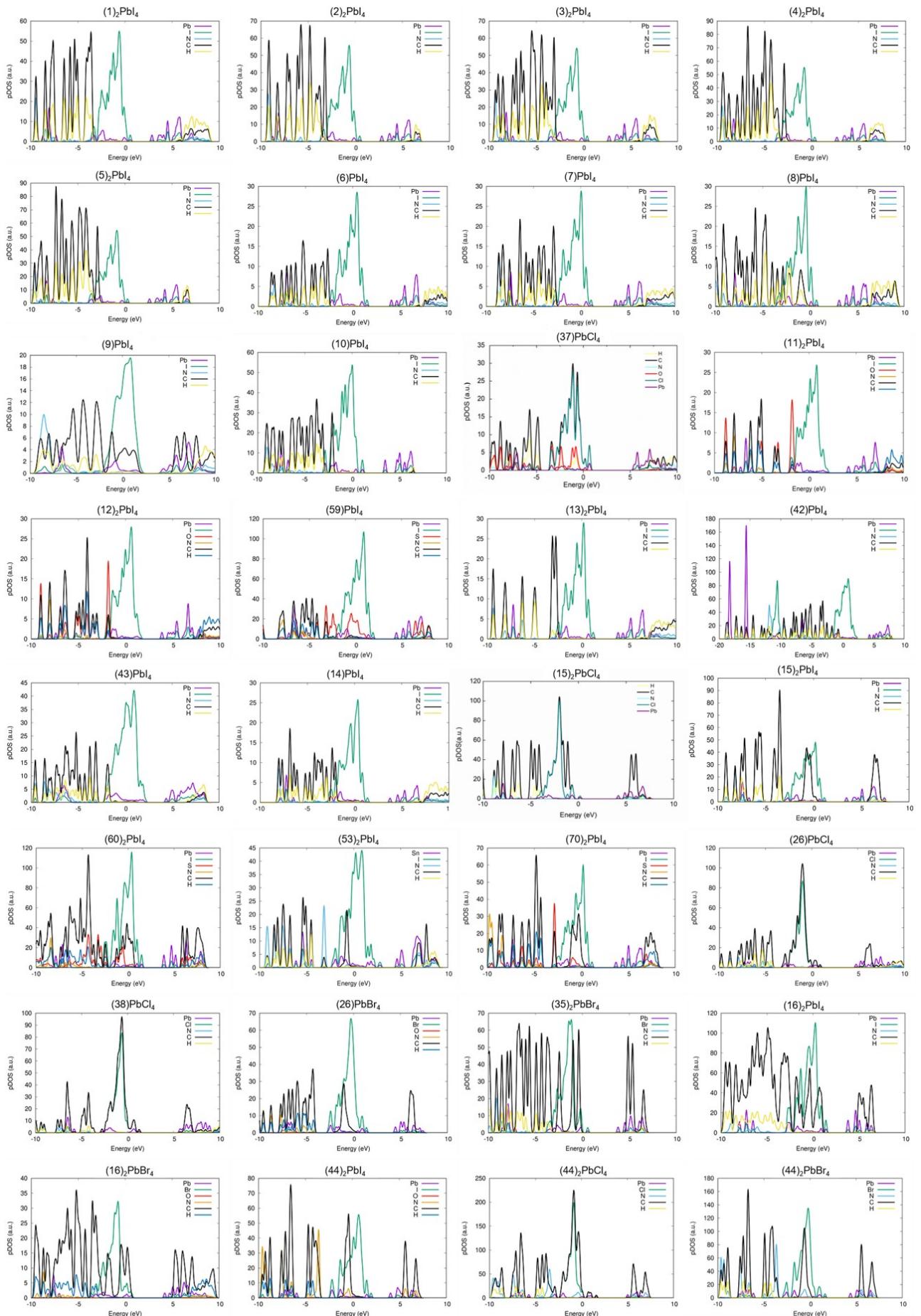
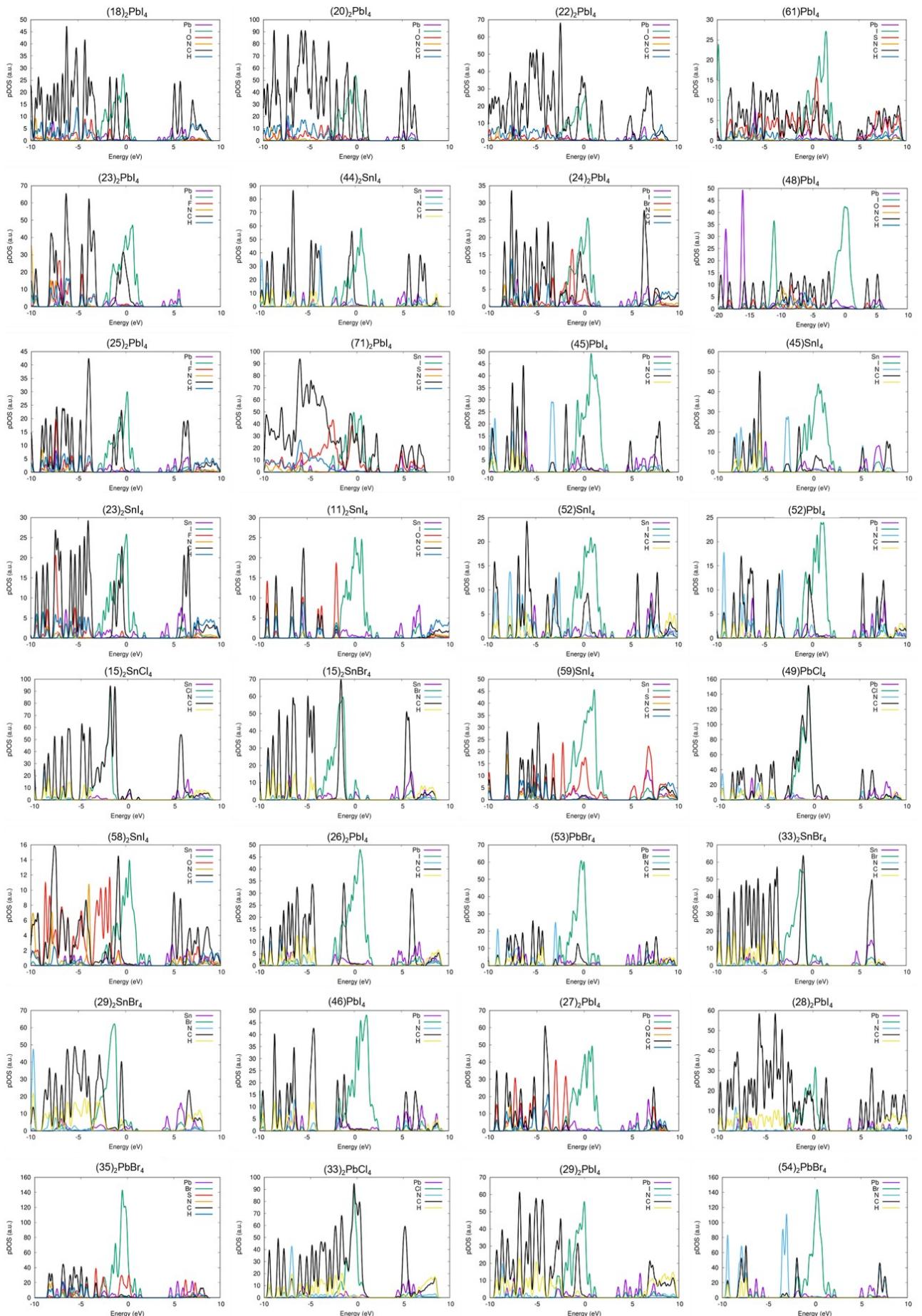


Figure S1: List of ammonium cations contained in the 2D hybrid perovskite materials selected in this study.





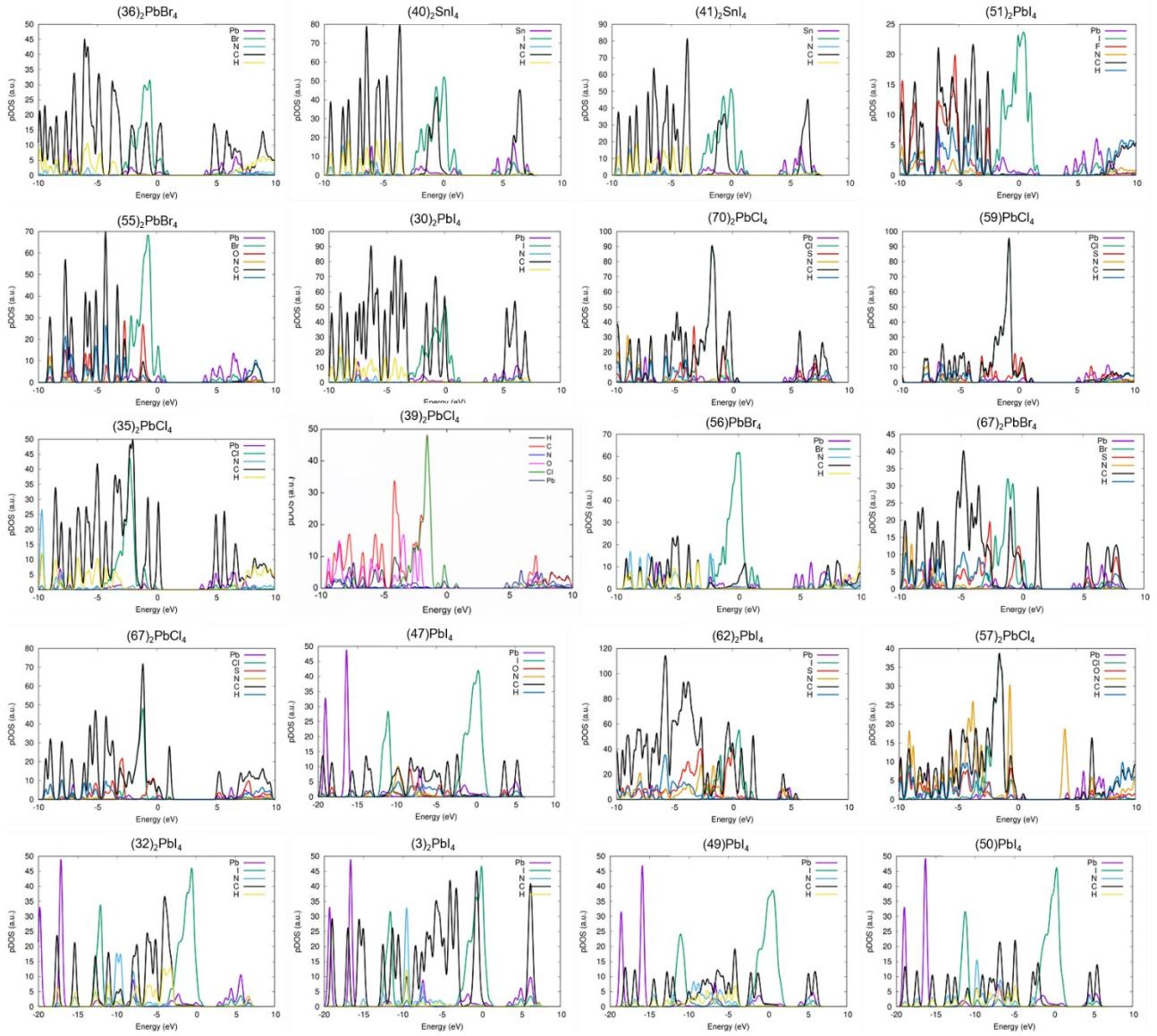


Figure S2: Projected density of states of 84 materials calculated by us. The structure of the numbered cation listed in the Figure S1.

Text S1: Computational details

To obtain the intramolecular band alignments of the layered hybrid perovskite materials we have performed electronic structure calculations using the Quantum Espresso (QE) program package.¹ Single point calculations at the HSE06+SOC level was performed. HSE06 hybrid functional with 43% Hatree-Fock exchange along with plane wave basis set cutoff of 40 Ry for the smooth part of the wave functions and a Fock energy cutoff of 80 Ry was set during HSE06+SOC calculations along with a Monkhorst pack k-point sampling.²⁻⁴ Also, we have taken norm-conserving pseudopotentials

with electrons from I 5s, 5p; Br 4s, 4p; Cl 3s, 3p; O, N, C 2s, 2p; H 1s; Pb 5s, 5p, 6s, 6p, 5d; Sn 4s, 4p, 5s, 5p, 4d; shells explicitly for the HSE06+SOC calculations. These methodologies are reported to give good results for such systems.⁵ After the single point calculations we have extracted the pDOS plots to identify the band alignment types. HOMO LUMO energies of the inorganic metal halide units were calculated using the same HSE06+SOC method. For this we have taken a single MX₆ unit neutralizing with four methyl ammonium cations. HOMO and LUMO energies of the six metal halide octahedral units were aligned with respect to the C 1S orbital energy.

For the calculation of HOMO LUMO energies of the isolated organic molecules the Gaussian 09 D.01 suite of programme has been employed.⁶ The Becke 3-parameter Lee–Yang–Parr (B3LYP)⁷⁻¹¹ hybrid functional has been implemented for the calculations. 6-311++G**¹²⁻¹⁴ basis set as formulated by John Pople was used.

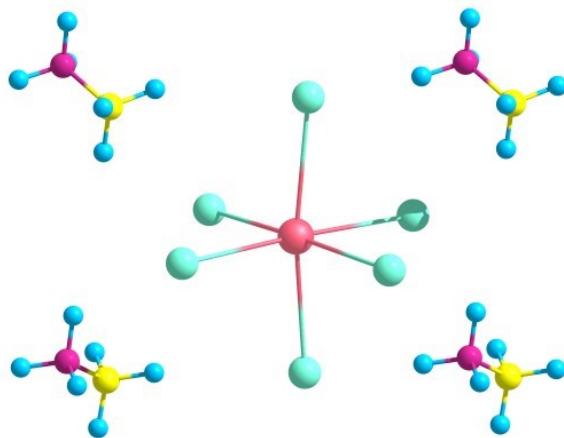


Figure S3: Single MX₆ unit with four methyl ammonium cations.

Text S2: Feature extraction details

We have taken the help of ChemAxon's Chemicalize platform to calculate the chemical descriptors for our work. To create the descriptors, we have drawn the structure of the ammonium cations there and calculated the descriptors. Among generated descriptors we have selected those features having direct influence on our output. In addition, another feature eccentricity which is the longest topological distance starting from the N atom inside the amine, we have calculated following the method described by Wu and co-workers¹⁵ in their recent paper.

Table S1: Selected features

Feature Name	Description
1. Organic HOMO	HOMO energy of the organic cation
2. Organic LUMO	LUMO energy of the organic cation
3. Inorganic HOMO	HOMO energy of the inorganic metal halide unit
4. Inorganic LUMO	LUMO energy of the inorganic metal halide unit
5. Metal Electronegativity	Electronegativity of the metal atom
6. Halogen Electronegativity	Electronegativity of the halogen atom
7. Eccentricity	Eccentricity of the organic cation
8. Molar Mass	Molar Mass of the organic cation
9. Ring Count	Number of rings present in the organic cation
10. Aromatic Ring Count	Number of aromatic rings present in the organic cation
11. Heterocyclic Ring Count	Number of heterocyclic rings present in the organic cation
12. Hydrogen Bond Donor Count	Number of functional groups having H-bond donors present in the organic cation
13. Hydrogen Bond Acceptor Count	Number of H-bond acceptor atoms present in the organic cation
14. Formal Charge	Formal Charge of the organic cation
15. Topological Polar Surface Area	Contribution of the polar atoms to the van der Waals surface area of the organic cation
16. Polarizability	Polarizability of the organic cation
17. Molar Refractivity	Molar Refractivity of the organic cation
18. van der Waals Volume	van der Waals volume of the organic cation
19. van der Waals Surface Area	van der Waals surface area of the organic cation
20. Minimum Projection Radius	Minimum radius of the organic cation when projected on a 2D plane

Text S3: Machine learning algorithms

Logistic Regression

Logistic regression is a linear model for classification. In this model, probabilities of the output of a single trial are demonstrated using a logistic function and it is also known as logit regression. This implementation can fit binary and multi-class regression with l1, l2, or elastic-net regularization.

Ridge Classifier

Ridge Classifier converts binary targets to {-1,1} and then treat the problem as regression task. Ridge regression addresses some of the problems of the ordinary least squares by imposing a penalty on the size of the coefficients. The ridge coefficients minimize a penalized residual sum of squares:

$$\min_w \|X_w - y\|_2^2 + \alpha \|w\|_2^2$$

The complexity parameter $\alpha \geq 0$ controls the amount of shrinkage: the larger the value of α , the greater the amount of shrinkage and thus the coefficients become more robust to collinearity.

Support Vector Machine (SVM)

Set of supervised learning methods such as regression, classification, outlier detection is included in SVM. In this work, we have used classification method. Here, the objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. To find the best hyperplane, various kernel functions can be used which helps to lead higher accuracy. Moreover, some other hyperparameters such as gamma, degree etc. also affect the accuracy. So, fine tuning of all these hyperparameters is very important. This method is also suitable for high dimensional dataset.

K-Nearest Neighbour (KNN)

KNN is a machine-learning algorithm that is used for classification and regression tasks. In case of classification, the algorithm works by finding the K nearest data points to a given test data point and then classifying them based on the majority class of these nearest neighbours. The value of k is a user-defined hyperparameter and can be chosen based on the problem and data at hand. The KNN algorithm can be formalized by calculating the distance between the test data point and all training data points. This can be done using Euclidean distance, Manhattan distance, or any other distance

metric. The overall steps can be written as following: (i) select the K nearest data points to the test data point based on the calculated distances. (ii) assign the majority class label to the test data point based on the class labels of the K nearest data points. If there is a tie, the algorithm can either choose a class at random or use some other method to break the tie.

Bagging Classifier

When random subsets of the dataset are drawn as random subsets of the samples, then this algorithm is known as Pasting. If samples are drawn with replacement, then the method is known as Bagging. A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction. Such a meta-estimator can typically be used as a way to reduce the variance of a black-box estimator (e.g., a decision tree), by introducing randomization into its construction procedure and then making an ensemble out of it. Here, we have used Random Forest as base classifier.

Random Forest Classifier

Random Forest Classifier is an extension of the Decision Tree algorithm that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification). In training, trees are grown using bootstrapped samples of the data and a random subset of the features. This results in a low correlation between the trees, reducing overfitting. During the prediction, the algorithm takes the average prediction across all the trees, providing more stability and robustness to outliers. Random Forest Classifier can handle high dimensional and non-linear data and is considered a robust algorithm for classification. However, it is computationally expensive and may have a longer training time compared to other algorithms.

Text S4: Repeated Stratified k-fold cross-validation

Stratified k-fold cross-validation is the strategic variation of popular k-fold cross-validation which returns stratified folds. This method is very useful for classification of an imbalanced dataset. During

the separation of training and test data using cross-validation, it keeps the ratio of two classes same in both the training and test set of each fold of the dataset. Repeated Stratified k-fold method repeats this whole process n times. In our case, we have a smaller number of type II materials and among type I materials, only 5 correspond to type Ib. Thus, here we have used Repeated Stratified k-fold cross-validation where 5 folds and 3 times repetition have been used. This method plays a very useful role to evaluate the classification metrics in this case.

Table S2: Predicted probability of test dataset (the numbers in the compound names are the numbers of the organic cations mentioned in Figure S1)

Compound	Type	Predicted probability	
		Type I	Type II
(22) ₂ PbI ₄	II	0.23	0.77
(71)PbCl ₄	II	0.24	0.76
(71)PbBr ₄	II	0.24	0.76
(48)PbI ₄	II	0.43	0.57
(63)PbI ₄	II	0.19	0.81
(63)PbBr ₄	II	0.19	0.81
(70) ₂ PbCl ₄	II	0.74	0.26
(4) ₂ PbI ₄	I	0.98	0.02
(5) ₂ PbI ₄	I	0.99	0.01
(6)PbI ₄	I	0.96	0.04
(59) ₂ PbI ₄	I	0.79	0.21
(13) ₂ PbI ₄	I	0.88	0.12
(42)PbI ₄	I	0.88	0.12
(24) ₂ PbI ₄	I	0.66	0.34
(25) ₂ PbI ₄	I	0.85	0.15
(45) ₂ PbI ₄	I	0.76	0.24
(36) ₂ PbBr ₄	I	0.66	0.34
(55) ₂ PbBr ₄	I	0.94	0.06
(33) ₂ PbI ₄	I	0.83	0.17
(34) ₂ PbI ₄	I	0.93	0.07
(71) ₂ SnI ₄	II	0.33	0.67
(15) ₂ SnCl ₄	II	0.78	0.22
(26)PbCl ₄	II	0.69	0.31
(53)SnI ₄	I	0.99	0.01

(41) ₂ SnI ₄	I	0.82	0.18
(15) ₂ SnBr ₄	I	0.78	0.22
(44) ₂ PbBr ₄	I	0.89	0.11
(77)PbCl ₄	I	0.88	0.12
(52)PbI ₄	I	0.47	0.53

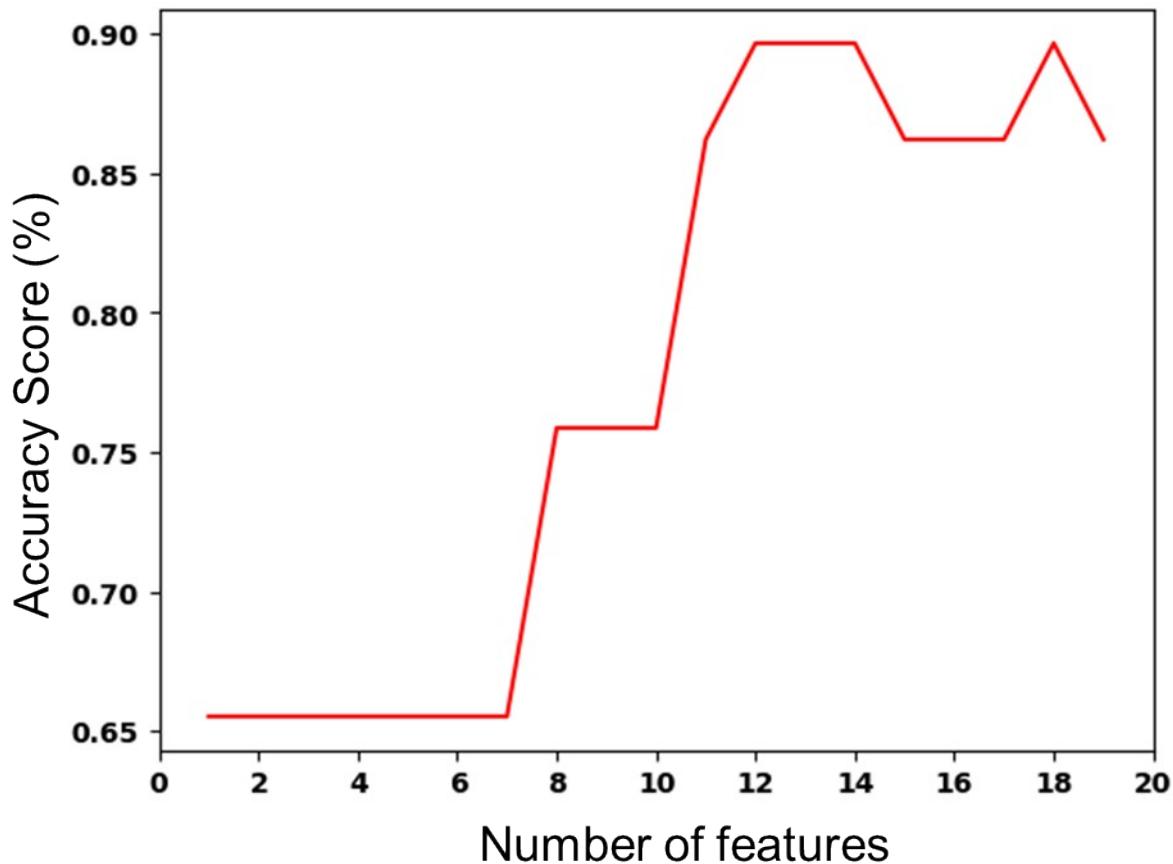


Figure S4: Accuracy score vs. number of features plot

Confusion Matrix		Predicted		$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
		Type I	Type II	
Actual	Type I	True Positive (TP)	False Negative (FN)	$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$
	Type II	False Positive (FP)	True negative (TN)	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$

F1 Score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

Figure S5: Confusion matrix and classification parameters

Text S5: Classification metrices

Each row in a confusion matrix represents an actual class, while each column represents a predicted class. Accuracy of a model is the number of correct predictions over total predictions in a binary classification problem. Precision is the exactness of a model that can be defined as among detected type I, how many are actually type I. Whereas Recall is the completeness of a model that tells us about among available type I, how many type I the model was able to detect correctly. A schematic representation and the formula of these metrices is presented in Figure S2.

Text S6: Model details

For the multiclass classification we used the same train and test sets that has been used in the I-II binary classification case. That is 74 instances in the train set and 29 instances in the test set. Whereas in Ia-Ib binary case we have separated the total 71 data to 50 in the train set and 21 in the test set. And in case of IIa-IIb binary classification where we have total 32 data instances divided them to 21 and 11 in the train and test sets respectively.

Table S3: Optimized hyperparameters and accuracy scores of binary classification between type I_a-I_b and type II_a-II_b and multi class classification between four types (I_a, I_b, II_a, II_b) using logistic regression.

Classification	Optimized Hyperparameters	Cross Validation (Average accuracy)	Test Accuracy
I _a -I _b	C: 0.01, penalty: l2, solver: liblinear	0.93	0.90
II _a -II _b	C: 100, penalty: l2, solver: liblinear	0.98	1.00
Multi class (I _a , I _b , II _a , II _b)	C: 1.0, penalty: l2, solver: liblinear	0.76	0.72

Table S4: Classification metrics for classification between I_a and I_b

Class	Precision	Recall	f1-score
type I _a	0.90	1.00	0.95
type I _b	0.00	0.00	0.00

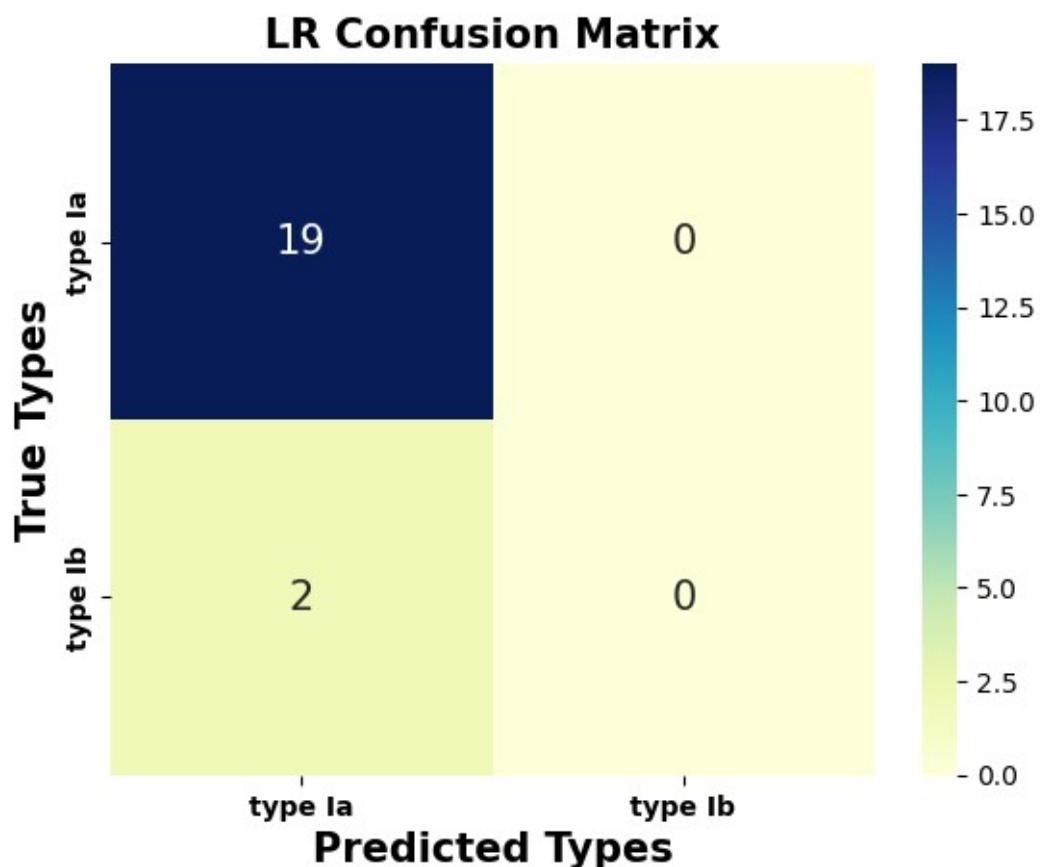


Figure S6: Confusion matrix for type I_a and type I_b classification of considered type I 2D perovskite.

Table S5: Classification metrics for classification between II_a and II_b

Class	Precision	Recall	f1-score
type II _a	1.00	1.00	1.00
type II _b	1.00	1.00	1.00

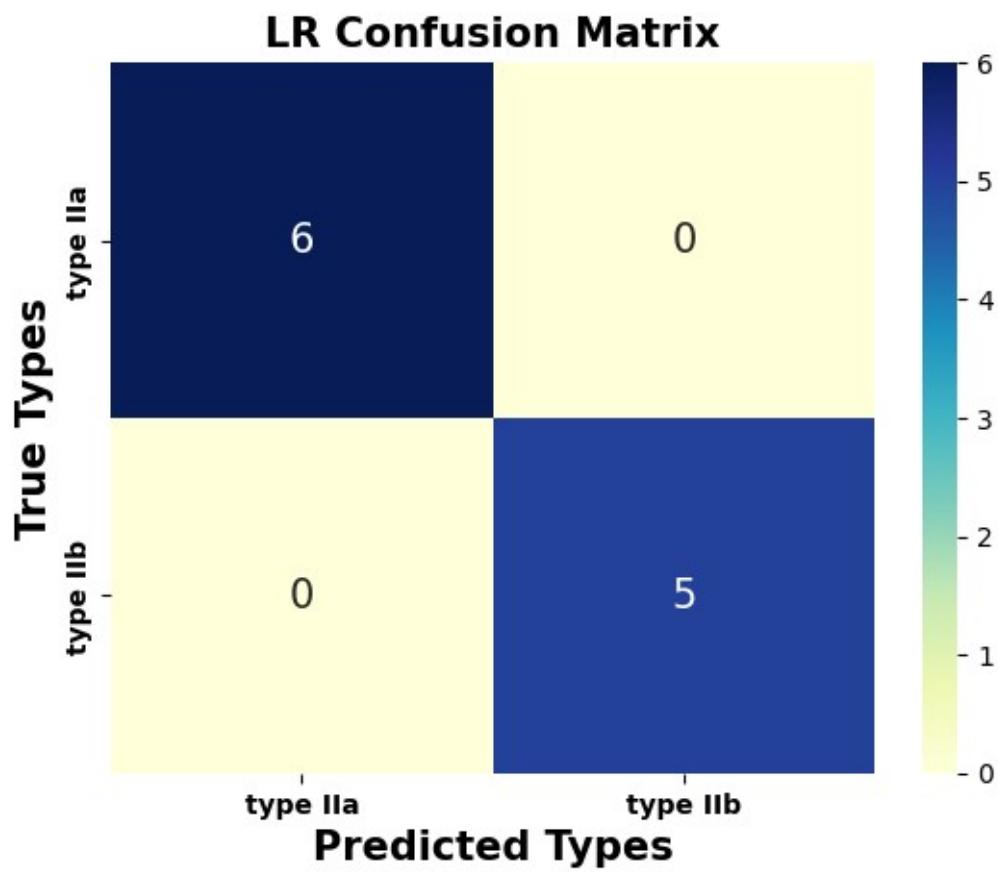


Figure S7: Confusion matrix for type II_a and type II_b classification of considered type II 2D perovskite.

Table S6: Classification metrics for multi class classification

Class	Precision	Recall	f1-score
type I _a	0.85	0.94	0.89
type I _b	0.00	0.00	0.00
type II _a	0.25	0.25	0.25
type II _b	0.60	0.50	0.55

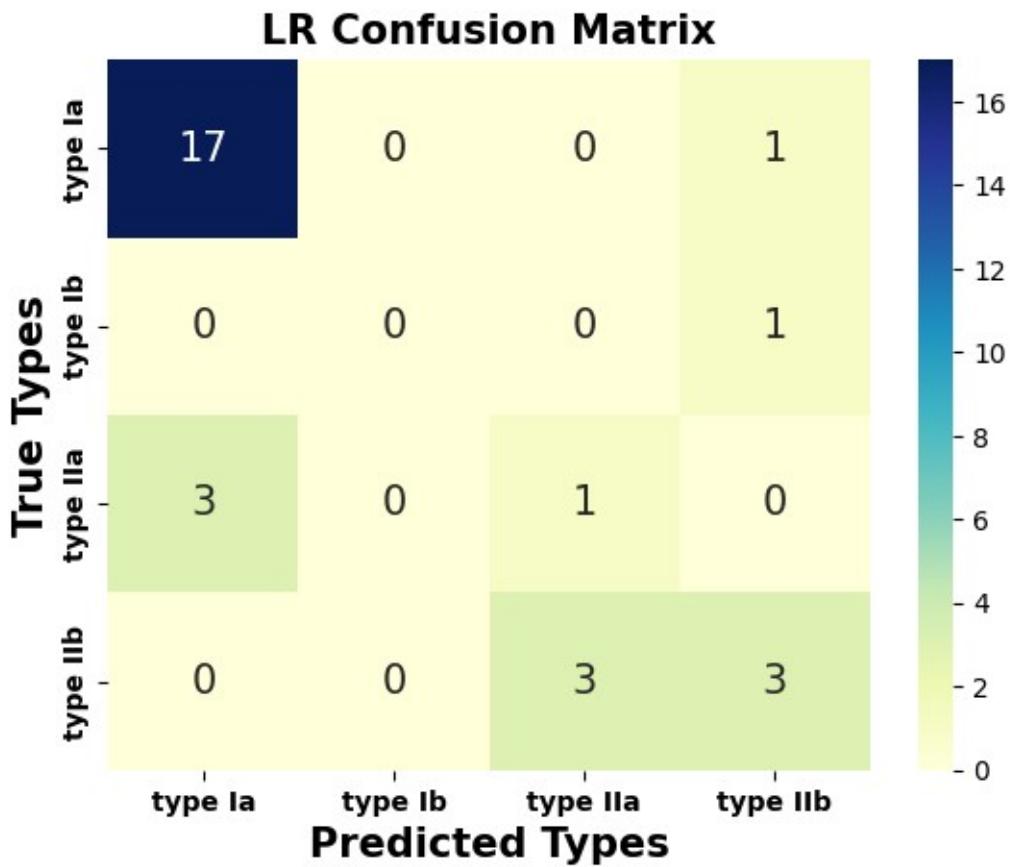


Figure S8: Confusion matrix for multi class classification of considered 2D perovskite.

Table S7: DFT validation of predicted unknown dataset (The feature names in the table are written in their abbreviated form mentioned in Table 2 in manuscript)

Compound	IH	IL	ME	HE	OH	OL	HB DC	HBA C	α	Actual type	Predicted type
[NH3C4H2S NH3]PbCl4	-3.8	-2.04	2.33	3.16	-16.072	-10.048	2	0	11.9	type II	type II
[NH3(CH2) C4H2S(CH2)NH3]PbCl4	-3.8	-2.04	2.33	3.16	-14.399	-8.8827	2	0	15.59	type II	type I
[C3H2SN(C H2)NH3]2P bBr4	-3.66	-2.06	2.33	2.96	-11.116	-5.3577	1	1	11.43	type I	type I
[NH3C4H2S NH3]PbBr4	-3.66	-2.06	2.33	2.96	-16.072	-10.048	2	0	11.9	type I	type I
[NH3C4H2S NH3]PbI4	-3.36	-2.06	2.33	2.66	-16.072	-10.048	2	0	11.9	type I	type I
[C2H4N3]2 PbI4	-3.36	-2.06	2.33	2.66	-14.173	-6.7604	2	1	6.55	type I	type I
[NH3C6H4 NH3]PbBr4	-3.66	-2.06	2.33	2.96	-16.004	-9.5268	2	0	13.2	type I	type I
[C3H3SN(C H2)NH3]PbI 4	-3.36	-2.06	2.33	2.66	-16.65	-11.35	2	0	11.69	type I	type I

[C3H2SN(C H2)NH3]2P bI4	-3.36	-2.06	2.33	2.66	-11.116	-5.3577	1	1	11.43	type I	type I
[H3NCH2C 10H6CH2N H3]PbI4	-3.36	-2.06	2.33	2.66	-12.7	-8.07	2	0	24.55	type I	type I
[CH3O(C2H 4)(CH2)2N H3]2PbI4	-3.36	-2.06	2.33	2.66	-9.773	-6.1177	1	6	17.14	type I	type I
[H3NCH2C 9H5NCH2N H3]PbI4	-3.36	-2.06	2.33	2.66	-13.43	-8.86	2	1	23.6	type I	type I
[H3NCH2C 8H4N2CH2 NH3]PbI4	-3.36	-2.06	2.33	2.66	-13.43	-8.86	2	2	22.68	type I	type I
[H3NCH2(C 4H2S)2CH2 NH3]PbI4	-3.36	-2.06	2.33	2.66	-12.1	-7.94	2	0	25.2	type I	type I
[C2H10S2N 2]PbI4	-3.36	-2.06	2.33	2.66	-15.39	-9.65	2	0	13.55	type I	type I
[(C4H2S)2C H2NH3]PbI 4	-3.36	-2.06	2.33	2.66	-8.95	-5.39	1	0	21.9	type II	type I
[H3NCH2(C 4H2S)2CH2 NH3]PbBr4	-3.66	-2.06	2.33	2.96	-12.1	-7.94	2	0	25.2	type II	type II

Table S8: Whole dataset of 103 2D perovskite materials considered in this work (Numbers in perovskite column are the number of materials in the 2D perovskite database developed by Tarasov and co-workers.¹⁷ The materials taken from literature are written with formula and the numbers in the formula are the numbers of the organic cations mentioned in Figure S1. The feature names in the table are written in their abbreviated form mentioned in Table 2 in manuscript).

Perovskite	IH	IL	ME	HE	OH	OL	Type	Multi class	HBDC	HBA C	α
19	-3.36	-2.06	2.33	2.66	-12.23	-5.48	type I	Ia	1	0	11.42
21	-3.36	-2.06	2.33	2.66	-11.59	-5.468	type I	Ia	1	0	13.27
23	-3.36	-2.06	2.33	2.66	-11.18	-5.459	type I	Ia	1	0	15.11
39	-3.36	-2.06	2.33	2.66	-14.52	-6.997	type I	Ia	2	0	18.41
40	-3.36	-2.06	2.33	2.66	-12.85	-6.465	type I	Ia	2	0	25.79
47	-3.36	-2.06	2.33	2.66	-14.22	-9.519	type I	Ia	2	0	20.82
53	-3.36	-2.06	2.33	2.66	-14.32	-6.88	type I	Ia	2	0	20.25
76	-3.36	-2.06	2.33	2.66	-12.85	-5.326	type I	Ia	2	1	6.51
77	-3.36	-2.06	2.33	2.66	-11.52	-5.446	type I	Ia	2	1	8.35
106	-3.36	-2.06	2.33	2.66	-10.08	-4.828	type I	Ia	2	0	13.98
111	-3.36	-2.06	2.33	2.66	-15.39	-7.486	type I	Ia	2	0	17.67
124	-3.36	-2.06	2.33	2.66	-10.93	-5.389	type I	Ia	1	0	13.66
127	-3.36	-2.06	2.33	2.66	-8.565	-5.161	type I	Ia	1	0	14.04
146	-3.36	-2.06	2.33	2.66	-9.284	-4.988	type I	Ia	1	0	23.06
152	-3.36	-2.06	2.33	2.66	-8.775	-5.163	type I	Ia	1	1	23.74

148	-3.36	-2.06	2.33	2.66	-11.3	-6.082	type I	Ia	2	0	14.96
153	-3.36	-2.06	2.33	2.66	-8.19	-5.448	type I	Ia	1	1	25.59
155	-3.36	-2.06	2.33	2.66	-7.958	-5.079	type II	IIa	1	1	35.58
156	-3.36	-2.06	2.33	2.66	-8.061	-4.6	type II	IIa	1	1	37.42
157	-3.36	-2.06	2.33	2.66	-8.086	-4.441	type II	IIa	1	1	39.26
159	-3.36	-2.06	2.33	2.66	-10.96	-6.939	type I	Ib	1	0	26.81
162	-3.36	-2.06	2.33	2.66	-10.92	-5.48	type I	Ia	1	0	13.28
338	-3.36	-2.06	2.33	2.66	-15.57	-9.157	type I	Ia	2	0	16.9
413	-3.36	-2.06	2.33	2.66	-15.91	-10.11	type I	Ia	1	0	14.75
439	-3.36	-2.06	2.33	2.66	-11.14	-5.652	type I	Ia	2	2	10.13
441	-3.36	-2.06	2.33	2.66	-7.742	-5.294	type II	IIa	1	0	38.63
456	-3.36	-2.06	2.33	2.66	-10.65	-5.056	type I	Ia	1	0	15.98
596	-3.36	-2.06	2.33	2.66	-9.089	-5.005	type I	Ia	1	0	26.53
687	-3.36	-2.06	2.33	2.66	-7.352	-4.848	type II	IIa	1	2	52.26
140	-3.66	-2.06	2.33	2.96	-15.57	-9.157	type I	Ia	2	0	16.9
145	-3.66	-2.06	2.33	2.96	-9.436	-5.283	type I	Ia	1	0	21.3
147	-3.66	-2.06	2.33	2.96	-9.284	-4.988	type I	Ia	1	0	23.06
351	-3.66	-2.06	2.33	2.96	-15.5	-8.813	type I	Ia	3	1	12.52
445	-3.66	-2.06	2.33	2.96	-14.24	-8.376	type I	Ia	2	0	17.19
532	-3.66	-2.06	2.33	2.96	-14.17	-6.76	type I	Ia	2	1	6.55
149	-3.8	-2.04	2.33	3.16	-11.3	-6.082	type II	IIa	2	0	14.96
618	-3.8	-2.04	2.33	3.16	-14.24	-8.376	type II	IIa	2	0	17.19
633	-3.8	-2.04	2.33	3.16	-9.436	-5.283	type II	IIa	1	0	21.3
675	-3.8	-2.04	2.33	3.16	-7.939	-5.46	type I	Ib	1	0	25.51
818	-3.36	-2.06	2.33	2.66	-8.381	-5.352	type II	IIa	1	0	34.83
(68) ₂ PbCl ₄	-3.8	-2.04	2.33	3.16	-8.746	-4.992	type II	IIa	1	0	23.66
(71)PbI ₄	-3.36	-2.06	2.33	2.66	-9.51	-6.326	type II	IIb	2	0	48
(71)SnBr ₄	-2.89	-2.02	1.96	2.96	-9.51	-6.326	type II	IIb	2	0	48
(71)SnI ₄	-3.17	-2.29	1.96	2.66	-9.51	-6.326	type I	Ib	2	0	48
(71)SnCl ₄	-3.26	-2.02	1.96	3.16	-9.51	-6.326	type I	Ia	2	0	48
(47)PbI ₄	-3.36	-2.06	2.33	2.66	-17.44	-11.92	type II	IIb	1	1	12.12
(64)PbI ₄	-3.36	-2.06	2.33	2.66	-16.97	-11.81	type II	IIb	1	0	14
(65)PbI ₄	-3.36	-2.06	2.33	2.66	-15.01	-10.56	type II	IIb	1	0	17.46
(66)PbI ₄	-3.36	-2.06	2.33	2.66	-15.91	-11.09	type II	IIb	1	0	15.61
(47)SnI ₄	-3.17	-2.29	1.96	2.66	-17.44	-11.92	type II	IIb	1	1	12.12
(64)PbBr ₄	-3.66	-2.06	2.33	2.96	-16.97	-11.81	type II	IIb	1	0	14
490	-3.36	-2.06	2.33	2.66	-16.81	-10.91	type I	Ia	2	0	12.96
494	-3.36	-2.06	2.33	2.66	-16.73	-11.09	type I	Ia	2	0	12.96
66	-3.8	-2.04	2.33	3.16	-14.17	-7.114	type I	Ia	2	2	16.11
112	-3.8	-2.04	2.33	3.16	-10.93	-5.389	type I	Ia	1	0	13.66
139	-3.8	-2.04	2.33	3.16	-16	-9.527	type II	IIa	2	0	13.2
646	-3.8	-2.04	2.33	3.16	-10.7	-5.58	type I	Ia	2	2	11.95
674	-3.66	-2.06	2.33	2.96	-7.939	-5.46	type II	IIa	1	0	25.51
699	-3.8	-2.04	2.33	3.16	-9.773	-6.118	type I	Ib	1	6	17.14
15	-3.36	-2.06	2.33	2.66	-12.84	-5.512	type I	Ia	1	0	9.57
325	-3.17	-2.29	1.96	2.66	-14.2	-8.38	type II	IIb	2	0	17.19
449	-3.8	-2.04	2.33	3.16	-10.6	-5.13	type II	IIa	1	0	15.43
337	-3.17	-2.29	1.96	2.66	-11	-6.43	type I	Ia	2	6	17.9
164	-3.17	-2.29	1.96	2.66	-11.3	-6.08	type I	Ia	2	0	14.96

269	-3.17	-2.29	1.96	2.66	-12.9	-5.33	type I	Ia	2	1	6.51
564	-3.17	-2.29	1.96	2.66	-10.8	-5.11	type I	Ia	1	0	15.51
356	-2.89	-2.02	1.96	2.96	-10.6	-5.13	type I	Ia	1	0	15.43
357	-2.89	-2.02	1.96	2.96	-10.6	-5.06	type I	Ia	1	0	15.98
289	-3.17	-2.29	1.96	2.66	-15	-10.1	type I	Ia	2	0	24.99
254	-3.17	-2.29	1.96	2.66	-10.5	-5.24	type I	Ia	1	0	15.04
652	-3.66	-2.06	2.33	2.96	-15.3	-9.18	type I	Ia	3	0	12.45
133	-3.36	-2.06	2.33	2.66	-10.7	-5.5	type I	Ia	1	0	12.28
567	-3.36	-2.06	2.33	2.66	-13.1	-5.2	type I	Ia	1	0	12.05
249	-3.17	-2.29	1.96	2.66	-15.8	-10.7	type I	Ia	4	0	14.08
158	-3.36	-2.06	2.33	2.66	-7.46	-5.03	type II	IIa	1	1	43.26
(71)PbCl ₄	-3.8	-2.04	2.33	3.16	-9.51	-6.33	type II	IIb	2	0	48
(71)PbBr ₄	-3.66	-2.06	2.33	2.96	-9.51	-6.33	type II	IIb	2	0	48
(48)PbI ₄	-3.36	-2.06	2.33	2.66	-17.3	-11.8	type II	IIb	1	1	12.12
(63)PbI ₄	-3.36	-2.06	2.33	2.66	-17	-11.7	type II	IIb	1	0	14.01
(63)PbBr ₄	-3.66	-2.06	2.33	2.96	-17	-11.7	type II	IIb	1	0	14.01
608	-3.8	-2.04	2.33	3.16	-10.7	-5.5	type II	IIa	1	0	12.28
26	-3.36	-2.06	2.33	2.66	-10.9	-5.45	type I	Ia	1	0	16.96
29	-3.36	-2.06	2.33	2.66	-10.5	-5.45	type I	Ia	1	0	18.81
38	-3.36	-2.06	2.33	2.66	-16.2	-7.54	type I	Ia	2	0	14.71
84	-3.36	-2.06	2.33	2.66	-14.2	-8.38	type I	Ia	2	0	17.19
86	-3.36	-2.06	2.33	2.66	-13.6	-5.5	type I	Ia	1	0	6.99
102	-3.36	-2.06	2.33	2.66	-17.2	-8.56	type I	Ia	2	0	13.98
175	-3.36	-2.06	2.33	2.66	-10.3	-5.48	type I	Ia	1	0	16.38
187	-3.36	-2.06	2.33	2.66	-10.5	-5.24	type I	Ia	1	0	15.04
248	-3.36	-2.06	2.33	2.66	-15.8	-10.7	type I	Ia	4	0	14.08
550	-3.66	-2.06	2.33	2.96	-9.7	-5.17	type I	Ia	1	0	23.14
592	-3.66	-2.06	2.33	2.96	-11.1	-5.36	type I	Ia	1	1	11.37
407	-3.36	-2.06	2.33	2.66	-10.6	-5.13	type I	Ia	1	0	15.43
(34) ₂ PbI ₄	-3.36	-2.06	2.33	2.66	-13.8	-5.77	type I	Ia	1	1	7.57
245	-3.17	-2.29	1.96	2.66	-7.2	-4.82	type II	IIb	1	0	46.45
302	-3.26	-2.02	1.96	3.16	-10.9	-5.39	type II	IIa	1	0	13.66
138	-3.8	-2.04	2.33	3.16	-15.6	-9.16	type II	IIa	2	0	16.9
129	-3.17	-2.29	1.96	2.66	-15.5	-8.81	type I	Ia	3	1	12.52
565	-3.17	-2.29	1.96	2.66	-10.8	-5.11	type I	Ia	1	0	15.51
303	-2.89	-2.02	1.96	2.96	-10.9	-5.39	type I	Ia	1	0	13.66
151	-3.66	-2.06	2.33	2.96	-11.3	-6.08	type I	Ia	2	0	14.96
331	-3.8	-2.04	2.33	3.16	-16.8	-11	type I	Ib	1	1	12.73
290	-3.36	-2.06	2.33	2.66	-15	-10.1	type I	Ia	2	0	24.99

References:

- (1) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gouguassis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.;

- Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: A Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.
- (2) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.
- (3) Du, M.-H. Density Functional Calculations of Native Defects in CH₃NH₃PbI₃: Effects of Spin–Orbit Coupling and Self-Interaction Error *J. Phys. Chem. Lett.* **2015**, *6*, 1461–1466.
- (4) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B: Solid State* **1976**, *13*, 5188–5192.
- (5) Mahata, A.; Mosconi, E.; Meggiolaro, D.; De Angelis, F. Modulating Band Alignment in Mixed Dimensionality 3D/2D Perovskites by Surface Termination Ligand Engineering. *Chem. Mater.* **2020**, *32*, 105–113.
- (6) Gaussian 09, Revision D.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, Jr., J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2013**.

- (7) Hehre, W. J.; Ditchfield, K.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (8) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (9) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (10) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (11) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (12) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- (13) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (14) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (15) Lyu, R.; Moore, C. E.; Liu, T.; Yu, Y.; Wu, Y. Predictive Design Model for Low-Dimensional Organic–Inorganic Halide Perovskites Assisted by Machine Learning. *J. Am. Chem. Soc.* **2021**, *143*, 12766–12776.
- (16) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *JMLR* **2011**, *12*, 2825–2830.
- (17) Marchenko, E. I.; Fateev, S. A.; Petrov, A. A.; Korolev, V. V.; Mitrofanov, A.; Petrov, A. V.; Goodilin, E. A.; Tarasov, A. B. Database of Two-Dimensional Hybrid Perovskite Materials: Open-Access Collection of Crystal Structures, Band Gaps, and Atomic Partial Charges Predicted by Machine Learning. *Chem. Mater.* **2020**, *32*, 7383–7388.

